



# BLASTing 16S RNA for Identification of Microbes

Searching 16S RNA sequences against curated entries from Targeted Loci for strain identification  
<https://blast.ncbi.nlm.nih.gov/>

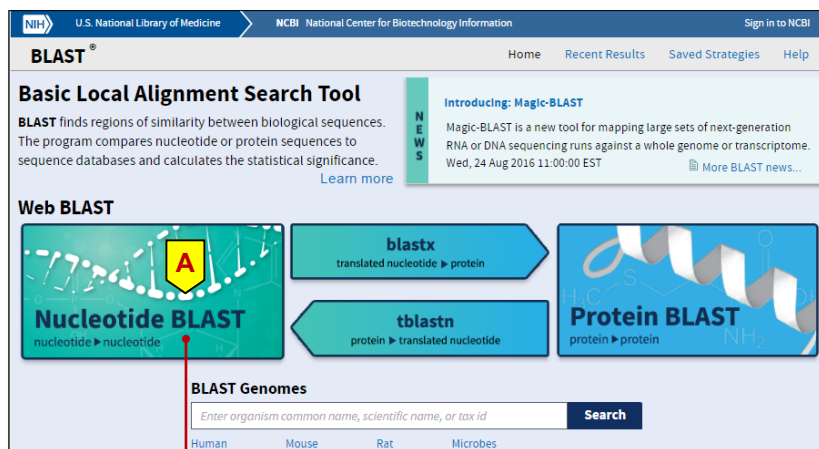
National Center for Biotechnology Information • National Library of Medicine • National Institutes of Health • Department of Health and Human Services

## Overview

Sequencing of highly conserved genes (also known as “targeted loci”), such as 16S ribosomal RNA and cytochrome C oxidase subunit I, found in samples collected from environmental and clinical specimens is crucial for identification of organisms and pathologic agents present in these samples, and important in phylogeny, population genetics, as well as microbial ecology studies. The BLAST web services [1] from NCBI offer a convenient tool to help you make sense of your 16S RNA sequence data. These include nucleotide BLAST (blastn) against the curated 16S Microbial ribosomal RNA database, or more broadly against the bacterial and archaeal subset in the NT database, and MOLE-BLAST [2], which performs a more comprehensive analysis and generates multiple alignment based phylogenetic tree(s) to provide a visual presentation of the relationships among the input query sequences in the context of matches database entries. MOLE-BLAST often provides good taxonomic placement for sequences that do not have perfect matches in the database.

## Setting Up the Search

From the NCBI BLAST homepage, click “nucleotide blast” link (A) under the “Basic BLAST” section to open the search page. Paste in the query sequences (B, provided as accessions), change the database from default “Nucleotide collection (nr/nt)” to “16S ribosomal RNA sequences (Bacteria and Archaea)” (C), and click BLAST button (D) to initiate the search.



NCBI/ BLAST/ blastn suite

### Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

Enter Query Sequence

BLASTN programs search nucleotide databases using a nucleotide query. [more...](#) [Reset page](#) [Bookmark](#)

Enter accession number(s), gi(s), or FASTA sequence(s)

AJ007639.1  
AJ007640.1  
AJ007643.1  
AJ007644.1  
AJ007645.1

Clear

Query subrange

From

To

Or, upload file

Choose File No file chosen

Job Title

AJ007639:Uncultured cyanobacterium 16S rRNA...

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database

☐ Human genomic + transcript ☐ Mouse genomic + transcript ☒ Others (nr etc.):

16S ribosomal RNA sequences (Bacteria and Archaea)

Organism

Optional

Enter organism name or id—completions will be suggested

Exclude

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

Limit to

Optional

Sequences from type material

Entrez Query

Optional

Enter an Entrez query to limit search

Program Selection

Optimize for

☒ Highly similar sequences (megablast)

☐ More dissimilar sequences (discontiguous megablast)

☐ Somewhat similar sequences (blastn)

Choose a BLAST algorithm

BLAST

Search database 16S ribosomal RNA sequences (Bacteria and Archaea) for highly similar sequences

☐ Show results in a new window

Algorithm parameters

Note: Parameter values that differ from the default are highlighted in yellow and marked with + sign

<https://go.usa.gov/xmeSx>

Genomic plus Transcript

Human genomic plus transcript (Human G+T)

Mouse genomic plus transcript (Mouse G+T)

Other Databases

Nucleotide collection (nr/nt)

16S ribosomal RNA sequences (Bacteria and Archaea)

Reference RNA sequences (refseq\_rna)

RefSeq Representative genomes (refseq\_representative\_genomes)

RefSeq Genome Database (refseq\_genomes)

Whole-genome shotgun contigs (wgs)

Expressed sequence tags (est)

Sequence Read Archive (SRA)

Transcriptome Shotgun Assembly (TSA)

High throughput genomic sequences (HTGS)

Patent sequences(pat)

Protein Data Bank (pdb)

Reference genomic sequences (refseq\_genomic)

Human RefSeqGene sequences(RefSeq\_Gene)

Genomic survey sequences (gss)

Sequence tagged sites (dbsts)

Such a BLAST search compares the input 16S RNA sequences from uncultured bacteria, already deposited in the Nucleotide database, against those 16S entries from the Targeted Loci projects curated by NCBI staff. The top identical or near identical match covering the whole length of the input query helps you identify the source organism of the input query sequences.

## 16S Ribosomal RNA BLAST Results

From the Description table, you can see that the second query matches to the 16S ribosomal RNA gene from *Synechococcus rubescens* strain with 100% length coverage and 99.70% identity (A, note the increased precision, note the increased precision). With this result, you can identify the query sequence to be from that organism. However, the result for the sixth query is less certain. Even though the query coverage is nearly complete at 99%, the identity is only 91.08% (B). Because of this lack of identical/near identical match, the source organism for this sequence cannot be ascertained. For this sequence, you can re-run the search using "Edit and Resubmit" link (C) to search against "(nr/nt)" database with uncultured/environmental sample sequences excluded, where a better match to an *Acinobacterium acl-B2* entry can be found.

**BLAST Results**

[Edit and Resubmit](#) [Search Strategies](#) [Formatting options](#) [Download](#)

Results for: 6.emb|AJ007640.1 Uncultured cyanobacterium 16S rRNA gene, clone 25/4-2, partial(845bp)

RID: FJWMDKG0014 (Expires on 06-08 04:07 am)

Query ID: AJ007640.1  
Description: Uncultured cyanobacterium 16S rRNA gene, clone 25/4-2, partial  
Molecule type: nucleic acid  
Query Length: 845  
Other reports: [Search Summary](#)

[Graphic Summary](#)  
[Descriptions](#)

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected: 0

[Alignments](#) [Download](#) [GenBank](#) [Graphics](#) [Distance tree of results](#)

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input type="checkbox"/> <i>Sporichthya polymorpha</i> DSM 43042 16S ribosomal RNA, partial sequence	1131	1131	99%	0.0	91.08%	NR_024727.1
<input type="checkbox"/> <i>Synechococcus rubescens</i> strain SAG 3.81 16S ribosomal RNA, partial sequence	1500		100%	0.0	99.70%	NR_125481.1

## The "16S Ribosomal RNA Sequences" Database

Sequences in this database are curated 16S ribosomal RNA sequences from two Targeted Loci projects, PRJNA33175 for bacteria (D) and PRJNA33317 for archaea (E). You can get the complete list of curation efforts from the umbrella project PRJNA224725 (<https://www.ncbi.nlm.nih.gov/bioproject/224725>).

**Archaea** Accession: PRJNA33317 ID: 33317

**Archaeal 16S Ribosomal RNA RefSeq Targeted Loci Project**

**Project Data:**

Resource Name	Number of Links
<b>SEQUENCE DATA</b>	
Nucleotide (total)	975
Genomic DNA	14
Transcript	961
<b>PUBLICATIONS</b>	
PubMed	451
PMC	97

<https://www.ncbi.nlm.nih.gov/bioproject/33317>  
(Only the data table is shown.)

**Bacteria** Accession: PRJNA33175 ID: 33175

**Bacterial 16S Ribosomal RNA RefSeq Targeted Loci Project**

The small subunit ribosomal RNA is a useful phylogenetic marker that has been used extensively for evolutionary analyses. The RefSeq dataset contains curated 16S ribosomal RNA sequences that correspond to bacteria type materials. The RefSeq records may contain corrections to the sequence or taxonomy as compared to the original INSD submission, and may have additional information added that is not found in the original. All sequences will have the same project ID and can be found as such.  
[NCBI RefSeq Targeted Loci Project](#). [Less...](#)

Accession	PRJNA33175
Data Type	RefSeq Targeted Locus (Loci) <a href="https://www.ncbi.nlm.nih.gov/bioproject/33175">https://www.ncbi.nlm.nih.gov/bioproject/33175</a>
Scope	Multisolate
Publications (total 11137)	1. López-Hermoso C <i>et al.</i> , "Salinivibrio kushneri sp. nov., a moderately halophilic bacterium isolated from salterns.", <i>Syst Appl Microbiol</i> , 2017 Dec 21;41(3):159-166 <a href="#">More...</a>
Submission	Registration date: 10-Dec-2008 <b>NCBI</b>
NCBI Links	• <b>Targeted Loci Project</b>
Related Resources	• <b>DSMZ</b> • <b>Silva</b> • <b>RDP</b> • <b>GreenGenes</b> • <b>StrainInfo</b> • <b>CRW</b>

**Project Data:**

Resource Name	Number of Links
<b>SEQUENCE DATA</b>	
Nucleotide (total)	20100
Genomic DNA	11
Transcript	20089
<b>PUBLICATIONS</b>	
PubMed	11146
PMC	1100

**NAVIGATE UP**

This project is a component of the RefSeq Targeted Loci Project

**NAVIGATE ACROSS**

9 additional projects are components of the RefSeq Targeted Loci Project

## Related Factsheets

- Guide to BLAST Web Site  
[https://ftp.ncbi.nlm.nih.gov/pub/factsheets/HowTo\\_BLASTGuide.pdf](https://ftp.ncbi.nlm.nih.gov/pub/factsheets/HowTo_BLASTGuide.pdf)
- MOLE-BLAST  
[https://ftp.ncbi.nlm.nih.gov/pub/factsheets/Factsheet\\_MOLE-BLAST.pdf](https://ftp.ncbi.nlm.nih.gov/pub/factsheets/Factsheet_MOLE-BLAST.pdf)